Multimodal Human Motion Capture and Synthesis

Marc Gowing[1], David S. Monaghan[1], Noel E. O'Connor[1]
Insight Centre for Data Analytics
marc.gowing2@mail.dcu.ie, david.monaghan@dcu.ie, noel.oconnor@dcu.ie

Abstract

Human motion capture (MoCap) involves the sensing, recording and mapping of human motion to a digital model. It is useful for many commercial applications and fields of study, including digital animation, virtual reality/gaming, biomechanical/clinical studies and sports performance analysis. What follows is an overview of the current research being carried out in Insight.

1. Motivation

Highly accurate motion capture systems are typically expensive and restricted to an indoor studio environment. The ability to capture motion in everyday surroundings using low cost equipment is highly desirable for several reasons, having potential for home rehabilitation and performance analysis of local/non-elite athletes in more realistic settings.

2. Problem Statement

Low cost depth sensors such as the Microsoft Kinect provide 2.5D scene geometry, greatly simplifying the task of foreground extraction and coarse pose recognition, crucial elements in MoCap. However, they are susceptible to occlusions and are generally unsuitable for large outdoor environments. Inertial sensors offer the most promising solution for both indoor and outdoor tracking, though they must be coupled with slower more accurate sensors/correction methods to avoid errors associated with drift. Additionally, as each body segment must have a sensor attached, this greatly increases the cost and intrusiveness of the system.

3. Related Work

Much research to date has focused on the use of cheap accelerometers or Inertial Measurement Units (IMU) attached to the body to estimate the orientation of body segments and infer full body pose [1]. To reduce the number of wearable sensors required, some authors [2] have attempted to fill in the information gaps by cross-referencing a pre-captured database of high quality MoCap data. The major limitation of such an approach is that the accuracy of the reconstructed motion depends on the richness/suitability of the database to the motions performed. A database such as this is typically recorded once using MoCap equipment and extensively post-processed, resulting in a static database that cannot be updated.

4. Research Question

Could depth and inertial sensors be combined to create a low cost human motion capture and synthesis system that attain accuracy comparable to expensive professional MoCap systems?

5. Hypothesis

This research will demonstrate that, the accuracy of pose estimation could be improved by fusing depth and inertial modalities for indoor environments. This motion capture system can in turn be used to create a personalized database of pre-captured motion, which can be used to synthesize motion from wearable sensors in more challenging environments.

6. Proposed Solution

Our low cost MoCap system is used to record a database of motions and construct a KD-tree for outdoor usage. The KD-tree is a data structure that partitions the database of available motions based on their similarity, facilitating efficient retrieval. During outdoor operation, for which depth sensors are not suitable, we rely on sparse inertial sensor data to query the KD-tree and synthesize full body pose.

In our approach, we overcome the bottleneck in database creation by using a Microsft Kinect to record the original full body motions and then subsequently enable the system to dynamically update and expand the database when unique motions are identified.

7. Evaluation

To evaluate the accuracy of the system, the motion synthesis results are compared to a Vicon system[2], a gold standard MoCap system. The accuracy of the system is presented in terms of the mean squared error for joint position. To date, using a database of Kinect motions and wearable accelerometers the system can achieve synthesis error of less than 14cm for the hands, and less than 5cm for joints closer to the body. This is comparable to Kinect skeleton tracking indoors, though the synthesis system is capable of working in both indoor and outdoor environments. Future research aims to reduce this error further using more sophisticated fusion techniques.

References

[1] D. Roetenberg. *Inertial and magnetic sensing of human motion*. University of Twente, 2006.
[2] R. Slyper and J. K. Hodgins. Action capture with accelerometers. In *Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 193–199. Eurographics Association, 2008.

[1] Acknowledgement - EU REVERIE FP7-ICT-287723
[2] www.vicon.com